Optimization Algorithms on Riemannian Manifolds and their Applications

小泉 孝弥

立命館大学理工学部数理科学科 4 回生

2021 年 3 月 15 日 (月曜日)
13:30∼14:30

Contents

Contents

最適化問題とは

最適化問題

M を $\mathbb{R}^n$ の部分集合とする．関数 $f : \mathbb{R}^n \to \mathbb{R}$ を M に制限した時に最大となる点 $x^* \in M$ を求める問題を最大化問題，最小となる点 $x_* \in M$ を求める問題を最小化問題という．最大化問題・最小化問題を合わせて (連続) 最適化問題 (Optimization Problem) といい，この時の f を目的関数 (objective function)，M を制約空間 (Constraint space) と呼ぶ．

特に，$M = \mathbb{R}^n$ の時，制約なし最適化問題といい，$M \neq \mathbb{R}^n$ の時，制約つき最適化問題という．また，f を $\mathbb{R}^n$ 上ではなく有限集合 $V \subset \mathbb{Z}$ 上の関数の時の最適化問題を (離散) 最適化問題と呼ぶ．

最適化問題の具体例

2 次関数の最小化 (制約なし)

$f : \mathbb{R} \to \mathbb{R}$ を $f(x) = x^2 + 25x + 12$ とする．この f が最小となるような点 $x^* \in \mathbb{R}$ を求めよ．

長方形の面積最大化 (制約あり)

$f : \mathbb{R}^2 \to \mathbb{R}$ を $f(x, y) = xy$ とする．$L \geq 0$ とした時，この f が最大となるような点
$(x^*, y^*) \in M = \{(x, y) \mid x \geq 0, y \geq 0, 2(x + y) = L\}$ を求めよ．

最適化数学と現実社会

現代社会において, 様々な分野で最適化は必要とされている.

Delivery Scope 問題

Covid-19 の影響で, 店舗配達の需要が増加し, 店舗の配達可能区域を決める必要性が高まった. できるだけ需要を高くした上で, 制限時間の間にお客に商品を届けることができる領域を求める問題.

高次元目的関数最適化問題

高次元 (何千億次元) で複雑な目的関数を最適化する問題 (特に, 機械学習).

Contents

最適化アルゴリズム

最適化問題を解くためにほぼ必須となるのが, **最適化アルゴリズム** (Optimization Algorithm) である. これにより, 最適化問題を計算機に効率的に解かせることができる. 例えば, 目的関数が微分可能で, 制約がない最適化問題であれば**勾配降下法** (Gradient Decent) を使って解くこと (局所的最適解を求める) ができる.

Algorithm 1 Gradient Decent(GD)

Require: f: differentiable function on $\mathbb{R}^d$
Require: $0 < t < 1$:
  $x \leftarrow x_0 \in \mathbb{R}^n$
  while x not converged **do**
    $x \leftarrow x - t \operatorname{grad} f(x)$
  end while
  return x

制約付き最適化問題とリーマン多様体

GD 法は便利な手法だが, $x \in M$ であったとしても, $x - t\, \mathrm{grad} f(x) \in M$ となるとは限らないため, GD 法を中心とした制約なし最適化問題用のアルゴリズムを制約付き最適化問題に用いることはできない．しかしながら, M がリーマン多様体である時は, f の定義域を M に制限することにより制約なしの最適化問題とみなすことで, これらのアルゴリズムをリーマン多様体に拡張したものを適用することができる.

Contents

Riemannian Manifold

Definition (Riemannian Manifold)

M を可微分多様体とする．任意の $x \in M$ の接空間 T_xM に内積 $g : T_xM \times T_xM \to \mathbb{R}$ が定まっている時, 組 (M, g) をリーマン多様体 *(Riemannian Manifold)* といい, g をリーマン計量 *(Riemannian metric)* と呼ぶ.

Definition (Gradient)

(M, g) をリーマン多様体とし, $f : M \to \mathbb{R}$ を可微分写像とする. $x \in M$ について

$$\forall \xi \in T_xM, g(\mathrm{grad} f(x), \xi) = Df(x)[\xi]$$

を満たす一意な $\mathrm{grad} f(x) \in T_xM$ を f の $x \in M$ での勾配 *(Gradient)* という.

Contents

具体例その1

Sphere

自然数 $n \geq 2$ について球面 $S^{n-1} := \{x \in \mathbb{R}^n \mid x^\top x = 1\}$ は $\mathbb{R}^n$ に埋め込まれた可微分多様体である．また, $x \in S^{n-1}$ での接空間 $T_x S^{n-1}$ は,

$$T_x S^{n-1} = \{z \in \mathbb{R}^n \mid x^\top z = 0\}$$

となる．$g_x : T_x S^{n-1} \times T_x S^{n-1} \to \mathbb{R}$ を

$$g_x(\xi, \eta) = \xi^\top \eta$$

と定めれば, g_x は内積となるので, S^{n-1} はリーマン多様体である．

Contents

具体例その2

Definition (Stiefel Manifold)

$X = [x_1 x_2 \dots x_p] \in \mathbb{R}^{n\times p} (n \geq p)$ で, $\{x_i\}_{i=1}^n$ が正規直交系であるような $n \times p$ 行列全体は $\mathbb{R}^{n\times p}$ に埋め込まれた可微分多様体となる．この多様体をシュティーフェル多様体 *(Stiefel Manifold)* といい, $\mathrm{St}(p, n)$ と表す．すなわち,

$$\mathrm{St}(p, n) = \left\{ X \in \mathbb{R}^{n\times p} \mid X^\top X = \mathbb{I}_p \right\}.$$

ただし, $\mathbb{I}_p$ は $p \times p$ の単位行列である．

$p = 1$ の時, $\mathrm{St}(1, n) = S^{n-1}$ であり, $n = p$ の時, $\mathrm{St}(n, n)$ は n 次の直交群 $O(n)$ となる．

シュティーフェル多様体の接空間

$X \in \mathrm{St}(p,n)$ での接空間 $T_X St(p,n)$ は,

$$T_X \mathrm{St}(p,n) = \left\{ Z \in \mathbb{R}^{n \times p} \mid X^T Z + Z^T X = 0 \right\}$$
$$= \left\{ X\Omega + X_\perp K \mid \Omega \in \mathrm{Skew}(p), K \in \mathbb{R}^{(n-p)\times p} \right\}$$

となる. ここで $\mathrm{Skew}(p) = \{\Omega \in \mathbb{R}^{p\times p} \mid \Omega^\top = -\Omega\}$ であり, $X_\perp$ は $\mathrm{span}(X)^\perp = \mathrm{span}(X_\perp)$ を満たす $n \times (n-p)$ 行列である.
$g_X : T_X \mathrm{St}(p,n) \times T_X \mathrm{St}(p,n) \to \mathbb{R}$ を,

$$g_X(\xi, \eta) = \mathrm{tr}(\xi^\top \eta)$$

と定めれば, g_X は内積となるので, $\mathrm{St}(p,n)$ はリーマン多様体である.

Contents

行列の最大・最小固有値問題とリーマン多様体

リーマン多様体上の最適化の具体例として，二次形式の最適化をすることで，対称行列の最大・最小固有値を求める問題を扱う．そのために，まず二次形式の導入を行い，重要な性質を紹介する．

Definition (二次形式)

$A \in \mathbb{R}^{n \times n}$ を対称行列とする．関数 $q_A : \mathbb{R}^n \to \mathbb{R}$ を

$$q_A(x) = x^\top A x$$

と定義する．この時，この関数 q_A を二次形式 *(quadratic form)* と呼ぶ．

二次形式と最大固有値・最小固有値

二次形式の重要な性質として, 以下の性質が知られている.

Proposition (二次形式と最大固有値・最小固有値 [2])

$A \in \mathbb{R}^{n\times n}$ を対称行列, $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ を A の固有値とする. この時,

1. $\forall x \in S^{n-1}, \lambda_1 \leq q_A(x) \leq \lambda_n$
2. $q_A(x_*) = \lambda_1$ となるのは, $x_* \in S^{n-1}$ が最小値 λ_1 の固有ベクトルであるのみである.
3. $q_A(x^*) = \lambda_n$ となるのは, $x^* \in S^{n-1}$ が最大値 λ_n の固有ベクトルであるのみである.

が成立する.

Optimization Problem on Sphere

先の性質に基づき、以下の球面上の最小化問題を考える.

球面上の二次形式最小化問題

$A \in \mathbb{R}^{n\times n}$ を対称行列とし, q_A を二次形式とする.

$$\left|\begin{array}{ll} \text{minimize} & q_A(x) = x^\top A x \\ \text{subject to} & x \in S^{n-1} \end{array}\right.$$

また, q_A を $-q_A$ とすれば最大固有値も求めることもできる.

Contents

勾配降下法の問題点

先の最適化問題を解くために, ユークリッド空間上の勾配降下法を球面上に拡張する. まず, ユークリッド勾配 $\mathrm{grad}_{\mathbb{R}^n} f(x)$ を, 球面上の勾配 $\mathrm{grad} f(x)$ に置き換えると, 以下のようになる.

1. 初期値 $x_0 \in S^{n-1}$ を決める.
2. $x \leftarrow x - t\ \mathrm{grad} f(x)$ として x を更新する.
3. 2 を収束するまで繰り返す.

しかし, $x - t\ \mathrm{grad} f(x) \notin S^{n-1}$ であるため, $\|x - t\ \mathrm{grad} f(x)\|$ で割り正規化することで, 更新後の点を S^{n-1} となるようにする.

球面上の勾配降下法

前スライドの議論を踏まえた上で, 球面上の勾配降下法を以下のように設計する.

Algorithm 2 Riemannian Gradient Decent on Sphere

Require: f: differentiable function on S^{n-1}
Require: $0 < t < 1$:
 $x \leftarrow x_0 \in S^{n-1}$
 while x not converged **do**
  $d = -t \,\mathrm{grad} f(x)$
  $x \leftarrow (x + d)/\|x + d\|$
 end while
 return x

球面上の勾配の求め方

Proposition (球面上の勾配と直交射影)

目的関数 $f : \mathbb{R}^n \to \mathbb{R}$ を S^{n-1} に制限した関数を $\overline{f} : S^{n-1} \to \mathbb{R}$ とする．この時, $x \in S^{n-1}$ について, $P_x : T_x\mathbb{R}^n \to T_xS^{n-1}$ を $T_x\mathbb{R}^n = \mathbb{R}^n$ から閉部分空間 T_xS^{n-1} への直交射影 *(orthogonal projection)* とすれば,

$$\mathrm{grad}\overline{f}(x) = P_x(\mathrm{grad}_{\mathbb{R}^n} f(x))$$

が成立する．また, 直交射影 P_x は

$$P_x(\xi) = (\mathbb{I}_n - xx^\top)\xi$$

で与えられる．

Contents

計算機実験

実際に具体的な対称行列 $A \in \mathbb{R}^{2\times 2}$ を考え，実際にその最小固有値に収束することを計算機を用いて確認する [3].

対称行列 $A \in \mathbb{R}^{2\times 2}$ を

$$A = \begin{pmatrix} 2 & 2 \\ 2 & 5 \end{pmatrix}$$

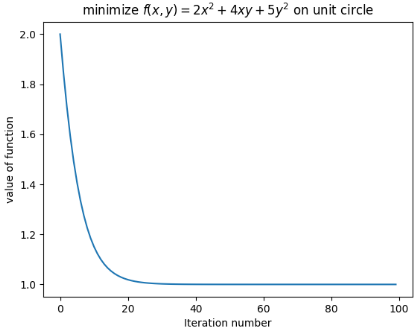
とする．この時，A の固有値は $\lambda_1 = 1, \lambda_2 = 6$ である．また，$q_A(x, y) = 2x^2 + 4xy + 5y^2$ より，

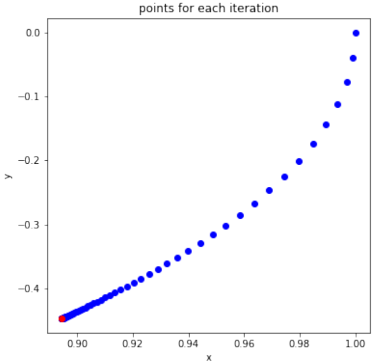
$$\mathrm{grad}_{\mathbb{R}^2} q_A(x, y) = (4(x + y), 4x + 10y)^\top$$

である．

実験結果 (最小固有値)

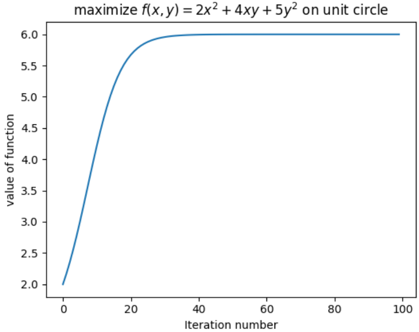
q_A の最小値 $\lambda_1 = 1$ に収束していることが確かめられた．また，その時の x は λ_1 の固有ベクトル $x_{\lambda_1} = (2/\sqrt{5}, -1/\sqrt{5})$ となっている．なお，$t = 0.01$, $x_0 = (1, 0)$ とした．

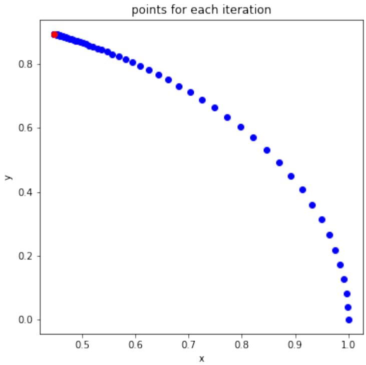




実験結果 (最大固有値)

q_A の最大値 $\lambda_2 = 6$ に収束していることが確かめられた．また，その時の x は λ_2 の固有ベクトル $x_{\lambda_2} = (1/\sqrt{5}, 2/\sqrt{5})$ となっている．なお，$t = 0.01$, $x_0 = (1, 0)$ とした．

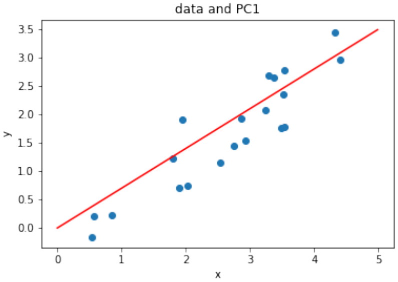




Contents

データの要約と主成分分析

主成分分析 (Principle Component Analysis, PCA) とは n 次元データ $\{x_i\}_{i=1}^N \subset \mathbb{R}^n$ をよく表現する正規ベクトル (大きさが 1 のベクトル) を見つける方法論. 簡単にいうと, 以下のような赤線を見つけること.



これ以降, データの平均は 0, 分散は 1 とする.

赤線の見つけ方

正規ベクトル v_1 が n 次元データ $\{x_i\}_{i=1}^N \subset \mathbb{R}^n$ を一番よく表現 (近似) するというのは, v_1 にデータを射影した時に, 射影前との誤差が最小になるということ．すなわち, $x \in \mathbb{R}^n$ の $\mathrm{span}\{v_1\}$ への直交射影は $v_1 v_1^\top x$ で与えられるから,

$$\mathop{\arg\min}_{\|v_1\|=1} \frac{1}{N} \sum_{i=1}^{N} \|x_i - v_1 v_1^\top x_i\|^2 \tag{1}$$

を求めること!.

標本共分散行列との繋がり

$\|x_i - v_1 v_1^\top x_i\|^2 = x_i^\top x_i - x_i^\top v_1 v_1^\top x_i$ だから，先ほどの (1) 式は以下と同値．

$$\mathop{\arg\max}_{\|v_1\|=1} \frac{1}{N}\sum_{i=1}^{N} x_i^\top v_1 v_1^\top x_i \tag{2}$$

ここで，データ $\{x_i\}_{i=1}^N$ の平均は 0 だから，$\{x_i\}_{i=1}^N$ の標本共分散行列 C は $C = \frac{1}{N}\sum_{i=1}^N x_i x_i^\top$ で与えられるので，

$$\mathop{\arg\max}_{\|v_1\|=1} \frac{1}{N}\sum_{i=1}^{N} x_i^\top v_1 v_1^\top x_i = \mathop{\arg\max}_{\|v_1\|=1} v_1^\top C v_1$$

となる．$\rightarrow$ C は対称行列だから，実はさっきやった問題と同じになる．

Contents

第 p 主成分

先ほどは, データ $\{x_i\}_{i=1}^N$ を一番近似する正規ベクトル v_1 を求めた. この v_1 を**第 1 主成分** (PC1) という. また, データを 2 番目によく近似し, v_1 と直交する正規ベクトル v_2 を**第 2 主成分** (PC2) という. 一般に, $v_1, v_2, \ldots, v_{p-1} (p < n)$ と直交し, データを p 番目によく近似する正規ベクトル v_k を**第 p 主成分** (PC p) という.

主成分分析

n 次元データが与えられた時, 上で定めたデータを近似する $1 \leq p \leq n$ 個の正規直交基底 $\{v_i\}_{i=1}^p$ を求めること.

p 次元への直交射影

前スライドより，主成分 $\{v_i\}_{i=1}^p$ を求めるためには以下の最適化問題を解けば良い．

$$\underset{v_1,v_2,\ldots,v_p:ONS}{\arg\min} \frac{1}{N}\sum_{i=1}^{N}\|x_i-(v_1v_1^T+\cdots+v_pv_p^T)x_i\|^2$$

ここで，$\{v_i\}_{i=1}^p$ は正規直交基底だから $(v_1v_1^T+\cdots+v_pv_p^T)x$ は $x\in\mathbb{R}^n$ の $\mathrm{span}\{v_1,v_2,\ldots,v_p\}$ への直交射影である
$(V=[v_1,v_2,\ldots,v_p]\in\mathbb{R}^{n\times p}$ とした時，$V(V^\top V)^{-1}V^\top=VV^\top)$．
また，$p=1$ の時と同様に考えれば上記の式は

$$\underset{v_1,v_2,\ldots,v_p:ONS}{\arg\max} \; v_1^\top Cv_1+v_2^\top Cv_2+\cdots+v_p^\top Cv_p$$

と同値．また，ここでトレースの定義より，以下のようにかける．

$$v_1^\top Cv_1+v_2^\top Cv_2+\cdots+v_p^\top Cv_p=\mathrm{tr}(V^\top CV)$$

シュティーフェル多様体と主成分分析

前スライドの議論と，シュティーフェル多様体が正規直交基底をなす $n \times p$ の集合であることを考えれば，主成分分析は以下のようなシュティーフェル多様体 $\mathrm{St}(p, n)$ 上の最適化問題と考えることができる．

C をデータの標本共分散行列とする．

$$\left| \begin{array}{ll} \text{minimize} & F(X) = -\operatorname{tr}(V^\top C V) \\ \text{subject to} & V \in \mathrm{St}(p, n) \end{array} \right.$$

Contents

最適化の準備：行列の極分解

球面上の二次形式最小化問題の時と同様に，勾配降下法をシュティーフェル多様体に拡張し解くことにする．しかし，これまた球面上に拡張した時と同様に $x - t\,\mathrm{grad} F(X) \notin \mathrm{St}(p, n)$ となる問題が発生する．今回は，これを解決するために行列の極分解 (polar decomposition) を利用する．

Proposition (行列の極分解)

任意の実行列 $A \in \mathbb{R}^{m \times n}$ は直交行列 U と半正定値対称行列 P を用いて

$$A = UP$$

と分解できる．

シュティーフェル多様体上の勾配降下法

Algorithm 3 Riemannian Gradient Decent on Stiefel Manifold

Require: F: differentiable function on $\mathrm{St}(p,n)$
Require: $0 < t < 1$:
  $X \leftarrow X_0 \in \mathrm{St}(p,n)$
  while X not converged **do**
    $d = -t\ \mathrm{grad}F(X)$
    $X \leftarrow (X+d)(\mathbb{I}_p + d^\top d)^{-1/2}$
  end while
  return X

なお, アルゴリズム中の $(X+d)(\mathbb{I}_p + d^\top d)^{-1/2}$ が, $X+d$ の極分解の U に対応している.

シュティーフェル多様体上の勾配の求め方

Proposition (シュティーフェル多様体の勾配と直交射影)

目的関数 $F : \mathbb{R}^{n\times p} \to \mathbb{R}$ を $\mathrm{St}(p,n)$ に制限した関数を $\overline{F} : \mathrm{St}(p,n) \to \mathbb{R}$ とする．この時，$X \in \mathrm{St}(p,n)$ について，$P_X : T_X\mathbb{R}^{n\times p} \to T_x\,\mathrm{St}(p,n)$ を $T_X\mathbb{R}^{n\times p} = \mathbb{R}^{n\times p}$ から閉部分空間 $T_X\,\mathrm{St}(p,n)$ への直交射影とすれば，

$$\mathrm{grad}\overline{F}(x) = P_X(\mathrm{grad}_{\mathbb{R}^{n\times p}}F(X))$$

が成立する．また，直交射影 P_X は

$$P_X(\xi) = \left(I - XX^T\right)\xi + X\,\mathrm{skew}\left(X^T\xi\right)$$

で与えられる．ただし，$\mathrm{skew}(A) = \frac{1}{2}(A - A^\top)$.

興味が出てきた人のために

1. 球やシュティーフェルの他にも, 実問題では Grassmann Manifold など他の多様体も登場する.
2. アルゴリズムに出てきた, 集合に戻すという写像をレトラクション (Retraction) という形で一般化される [1].
3. 最近, 一部のリーマン多様体においてレトラクションがいらないという話が出てきた [5].

References

[1] P.-A. ABSIL, R. MAHONY, AND R. SEPULCHRE, Optimization Algorithms on Matrix Manifolds, Princeton University Press, Princeton, NJ, 2008.

[2] https://www.cck.dendai.ac.jp/math/support/advance/2 次形式.pdf, 東京電機大学

[3] 佐藤 寛之, 曲がった空間での最適化, 日本オペレーションズ・リサーチ学会, オペレーションズ・リサーチ Vol.60 9 月号, 2015

[4] 佐藤一宏, 線形数理要論 (第 8 回), `http://www.kazuhirosato.work/entry/senkeisuriyoron_2020`, 2020

[5] Pierre Ablin and Gabriel Peyré, Fast and accurate optimization on the orthogonal manifold without retraction, `https://arxiv.org/abs/2102.07432`, 2021